Data Analysis and Statistics SIG Meeting Minutes

|	
Date, Time & Location:	September 3, 2004 2:00 – 3:00 PM EDT
Attendees:	Simon Lin – Duke Bob Clarke - Georgetown Steve Marron – UNC-Lineberger Tom Moloshok – Moffit Julie Zhu – Northwestern Michael Showe - Wistar Michael Nebozhyn - Wistar Claire Zhu – BAH Juli Klemm - BAH
Introduction:	**Roll-call, open meeting, review meeting goals** - Review discussions from last meeting - Discussion of use case request and example from Architecture Workspace - Review discussions from face-to-face meeting and discuss next steps for the SIG - Set agenda for next meeting
Review Discussions:	**Review discussions from last meeting** - The group had presentations of DWD (UNC) and VISDA (Georgetown). **Discussion of use case request and example from Architecture Workspace** - The Architecture Workspace requested submission of use cases from ICR SIG. The purpose is to help the Architecture Workspace define architectural issues that are relevant to this domain workspace. - Patrick McConnell from Duke, who is the ICR liaison to the Architecture Workspace, has put together an example use case based on the RProteomics project, which is one of the grid reference implementation projects in caBIG. - In Patrick's example, 3 levels of use cases were presented: 1, how the data are used now, 2, how the data will be used at the end of the one year project, and 3, how the data will be used in a grid like environment. - Juli has distributed Patrick's example along with the meeting agenda. - Steve Marron pointed out that the use cases should focus on how others would use the data and tools, not developers themselves. - Timeline for submission is Friday, 9/17, approximately 2 weeks from today. - The use cases to be submitted to the Architecture group should be different from the formal use cases that the Adopters draw for Developers for workspace projects. The purpose is to provide useful generic information for the Architecture group, and there is no commitment involved. **Review discussions from face-to-face meeting and discuss next steps for the SIG** - One important request came out of the face-to-face meeting is support and guidance from cross cutting workspaces at the earliest stages of project development. Juli has been talking with cross cutting workspace leads on the
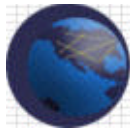
possible models of interaction.

- Given the large number of projects within ICR, it is impractical to have one-on-one interactions. Therefore, Juli suggested to have one point person assigned to each SIG. Concerns about this model that were raised at other SIG teleconferences include over-burdening of the point person, response time, etc.

- How many projects are there in this SIG? (Steve M)

    o There are 5 projects in the data analysis SIG.

- Where to find project information within ICR? (Julie Zhu)

    o A spreadsheet containing all projects, descriptions and Developer-Adopter pairs will be distributed in the next few days. This information has been very dynamic, and projects have been finalized only recently.

- What is the appropriate turn-around time for responses? (Steve M)

    o It depends on the project, and is hard to determine ahead of time. The matrix for any model of interaction should be quality of information and turn-around time.

- What are the backgrounds of the people in the Architecture group? (Steve M)

    o The Architecture group consists of members from NCICB, its contractors, BAH, and the Cancer Centers.

- In terms of coordination, another way of doing it is to put people in groups of specific skills and have them provide answers to questions through emails. (Steve M)

    o A related point is that there is Listserv for a number of caCORE infrastructures, e.g. caBIO developer listserv, caBIO user listserv, etc. They are well run and responsive. They are available on NCICB websites.

- Maybe an ideal model would be to integrate listserv and point of contacts, i.e. if one has a question, one would post it on a listserv, at the same time also bring it to the point of contact from the Architecture group. (Steve M)

    o Good model. It puts a moderator on the requests. Will discuss it with the cross cutting workspace leads.

- Juli suggested appointing a SIG lead for coordination across SIGs. The role of the SIG lead is to identify issues that are relevant to all workspace projects, and to coordinate the follow ups on these issues. Juli wanted the group's opinions on this, and whether there is the need for such a role.

- Steve Marron thinks that there is a need for this, and it should be a dedicated job, maybe bringing a new person in for it.

- Tom Moloshok prefers the self-organizing mode, in which the project teams directly deal with the issues themselves.

Face-to-face meeting follow-up discussion: Description of data analysis methods

- Another important issue raised at the face-to-face meeting is the need to capture data analysis methods. Recording how raw data has been transformed and what manipulations have been done is very important for reproducing analysis results.

- There need to be guideline on data analysis procedure. (Bob Clark)
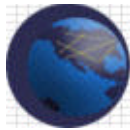
	- Simon Lin speculated on what it takes to capture data analysis processes: 1, write done exactly what you did in script, 2, document the platform on which the codes were run, 3, document the machine on which the codes were run. The problem seems to be the lack of cross-language capability. There is a need for ontology to describe procedure names. - Anyone aware of any existing model for tracking data analysis? (Juli) ○ The Statistical Metadata Standard from the World Bank. (Simon) - This is an issue to bring up to the cross-cutting workspaces

Other Items Discussed:	- CaBIG forum
	
	    ○ As workspace projects are starting to kick off, the forum can be a very effective tool for discussions of project or SIG specific issues.
	
	    ○ Juli encouraged everyone to explore the forum more!
	
	    ○ A username and a password are required to post on the forum. If you don't already have them, let Juli know.
	
	    ○ Juli will start using the forum herself by posting meeting materials.
	
	    ○ The caBIG web site has been re-designed. It will be easier to access to the forum.

Action Items:	Name Responsible	Action Item	Date Due	Notes
	Juli Klemm	Post materials on the forum		
	Juli Klemm	Distribute meeting minutes		